Pearce KF, Balavarca Y, Norden J, Jackson G, Holler E, Dressel R, Greinix H, Toubert A, Gluckman E, Hromadnikova I, Sedlacek P, Wolff D, Holtick U, Bickeboller H, Dickinson AM.

Impact of genomic risk factors on survival after haematopoietic stem cell transplantation for patients with acute leukaemia.

***International Journal of Immunogenetics* 2016, 43(6), 404-412.**

Impact of genomic risk factors on survival after haematopoietic stem cell transplantation for patients with acute leukaemia.

Disclosures

The authors declare no conflict of interest.

Funding

This work was supported by the Marie Curie Research Training Network (MCRTN) grant CT-2004-512253: TRANSNET (European Commission), grant LSHB-CT-2007-037703: Stemdiagnostics (European Commission), the German Research Foundation (DFG), grant GRK 1034, the Marie Curie Initial Training Network (MCITN) grant 315963: CELLEUROPE and the Deutsche José Carreras Leukämie-Stiftung.

Kim F. Pearce[1], Yesilda Balavarca[2], Jean Norden[1], Graham Jackson[3], Ernst Holler[4], Ralf Dressel[5], Hildegard Greinix[6], Antoine Toubert[7], Eliane Gluckman[8], Ilona Hromadnikova[9], Petr Sedlacek[10], Daniel Wolff[4], Udo Holtick[11], Heike Bickeböller[2] and Anne M. Dickinson[1]

Email addresses: Kim.Pearce@ncl.ac.uk, yesilda.balavarca@nct-heidelberg.de, jean.norden@ncl.ac.uk, graham.jackson@ncl.ac.uk, ernst.holler@Klinik.uni-regensburg.de, rdresse@gwdg.de, Hildegard.Greinix@klinikum-graz.at, antoine.toubert@univ-paris-diderot.fr, eliane.gluckman@sls.aphp.fr, ilona.hromadnikova@lf3.cuni.cz, petr.sedlacek@lfmotol.cuni.cz, daniel.wolff@klinik.uni-regensburg.de, udo.holtick@uk-koeln.de, hbickeb@gwdg.de, anne.dickinson@newcastle.ac.uk.

[1]Haematological Sciences, Institute of Cellular Medicine, Newcastle University, UK; [2]Department of Genetic Epidemiology, University Medical Center, Göttingen, Germany; [3]Northern Centre for Cancer Care, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK;

[4]Department of Internal Medicine III, University of Regensburg, Regensburg, Germany;

[5]Department of Cellular and Molecular Immunology, University Medical Center, Göttingen, Germany;

[6]Department of Internal Medicine, Division of Haematology, Medical University of Graz, Graz, Austria;

[7]Departement d′Immunologie, Université Paris Diderot, INSERM UMRS-940, AP-HP, Paris, France;

[8]Department of Bone Marrow Transplantation, EUROCORD, St Louis Hospital, Paris, France;

[9]Department of Molecular Biology and Cell Pathology, Third Faculty of Medicine, Charles University Prague, Prague, Czech Republic;

[10]Department of Pediatric Hematology and Oncology, Second Faculty of Medicine, Charles University Prague, Prague, Czech Republic;

[11]Department I of Internal Medicine, University of Cologne, Cologne, Germany;

HB and AMD contributed equally to the work and are joint last authors

Corresponding author (address for proofs and reprints): Dr Kim F. Pearce. Haematological Sciences, Institute of Cellular Medicine, Newcastle University, UK. Email: Kim.Pearce@ncl.ac.uk

Summary

The EBMT risk score is an established tool successfully used in the prognosis of survival post HSCT and is applicable for a range of haematological disorders. One of its main advantages is that score generation involves summation of clinical parameters that are available pre-transplant. However, the EBMT risk score is recognised as not being optimal.

Previous analyses, involving patients with various diagnoses, have shown that non-HLA gene polymorphisms influence outcome after allogeneic HSCT. This study is novel as it focuses only on patients having acute leukaemia (N=458) and attempts to demonstrate how non-HLA gene polymorphisms can be added to the EBMT risk score in a Cox regression model to improve prognostic ability for overall survival. The results of the study found that three genetic factors improved EBMT risk score. The presence of *MAL* (rs8177374) allele T in the patient, absence of glucocorticoid receptor haplotype (consisting of rs6198, rs33389 and rs33388) ACT in the patient and absence of heat shock protein 70-hom (+2437) (rs2227956) allele C in the patient were associated with decreased survival time. When compared to the EBMT risk score, the scores combining EBMT risk score with the genetic factors had an improved correlation with clinical outcome and better separation of risk groups. A bootstrapping technique, involving repeated testing of a model using multiple validation sets, also revealed that the newly proposed model had improved predictive value when compared to the EBMT risk score alone.

Results support the view that non-HLA polymorphisms could be useful for pre-transplant clinical assessment and provide evidence that polymorphisms in the recipient genotype may influence incoming donor cells, suppressing the initiation of the graft versus leukaemia effect and reducing survival.

Introduction

Allogeneic haematopoietic stem cell transplantation (HSCT) is the principal curative therapy for disorders of the haematopoietic system but it is still associated with substantial morbidity and mortality (Copelan (2006), Gratwohl *et al.* (2007)) with a survival rate of 40-60% (Balavarca *et al.* (2015)). Five main clinical factors influencing outcome after HSCT have been established by the European Group for Blood and Marrow Transplantation (EBMT). These EBMT-factors (patient age, stage of the disease, time interval from diagnosis to transplant, histocompatibility and donor and patient gender combination) have been successfully utilised in a clinical risk score (EBMT risk score) (Gratwohl (2012)) for all patients undergoing HSCT for a haematological disorder (Gratwohl *et al.* (2009)) (Supplementary Table 1).

Recently several groups, including ourselves, have used both small cohorts (Dickinson (2008), Mullally & Ritz (2007)) and genome-wide association studies (GWAS) (Chien *et al.* (2012)) to demonstrate that polymorphisms within non-HLA genes are predictive of HSCT outcome. The effect of gene polymorphisms may be influenced by year of transplant; transplant protocols have been modified over the last ten years due to increased use of reduced intensity conditioning, increased use of peripheral blood stem cells rather than bone marrow and increased age of the patient at time of transplant (Feinstein *et al.* (2001)) thus altering the biology of HSCT.

In this study we investigated the role of non-HLA polymorphisms in a cohort of acute leukaemia patients with HLA matched sibling and matched unrelated donors. The influence of change in transplant protocols over time was also assessed. The results from the non-HLA genetic analysis were used to derive a novel (and validated) statistical model for overall survival which integrated non-HLA genetics into the EBMT risk score. Gene polymorphisms

were also investigated for their association with non-relapse mortality, relapse and acute and chronic graft versus host disease.

All patients and donors gave informed consent in accordance with EBMT guidelines and the Declaration of Helsinki. The protocol was approved by the Local Research Ethics Committee at the co-ordinating centre (Newcastle upon Tyne).

Materials and Methods

Patients

Data were available from a total of 458 adult acute leukaemia patients (330 acute myeloid leukaemia, 123 acute lymphoblastic leukaemia, 3 acute biphenotypic leukaemia and 2 acute undifferentiated leukaemia) and were collected prospectively from 8 European transplant centres. Transplants were conducted between 1983 and 2005. All clinical data were collected from the EBMT database ProMIse. In this cohort, 58% of patients had died and 30% of patients had relapsed. HLA typing was determined by standard technologies using serology or molecular typing. Of the 458 patients, 29% received reduced intensity conditioning and 34% were T cell depleted. All patients received standard comparative supportive care across centres. The median follow up time was 77 months. Causes of death were stated as relapse (46%), acute respiratory distress syndrome (ARDS)/infection (21%), graft versus host disease (GvHD) (18%) and 'other' (15%). Clinical relapse was defined as hematologic, cytogenetic or molecular relapse (Branford (2007)). Patient and donor characteristics are presented in Table 1.

Genotyping for cytokine polymorphisms

Archived frozen peripheral blood mononuclear cells were used for the preparation of DNA. Genotyping was carried out by Kbioscience (http://www.kbioscience.co.uk) who used fluorescence-based competitive PCR technology (KASPar) and designed the assays for the SNPs based on the DNA sequence (50 bases) either side of the SNP. Genotypes were

available for 32 non-HLA candidate polymorphisms from 20 genes (Supplementary Table 2). These SNPs were chosen according to findings on smaller patient cohorts by our coordinating centre in Newcastle (*ESR1* (Middleton *et al.* (2003)), *IFNG* (Pravica *et al.* (1999)), *IL1RN* (Hurme & Santtila (1998)), *IL4* (Walley & Cookson (1996)), *IL6* (Cavet *et al.* (2001),Yee *et al.* (2009)), *IL10* (Morse *et al.* (1999)), *IL13* (Graves *et al.* (2000)), *TNF* (Imboden *et al.* (2006)), *TNFRSF1B* (Stark *et al.* (2003)) and *VDR* (Middleton *et al.* (2002))) and according to findings by other HSCT groups (*MAL* (Rocha *et al.* (2007)), *MDR1* (Kim *et al.* (2006)) and *NOD2* (Holler *et al.* (2004)). Candidate SNPs were also selected according to previous disease association studies in autoimmune (*CD14* (Klein *et al.* (2002)), GCR (Derijk *et al.* (2001), Stevens *et al.* (2004)), HSP70-hom (Bogunia-Kubik & Lange (2005)) and *IL12B* (Cargill *et al.* (2007))) or inflammatory disease and GvHD (Novota *et al.* (2011)) (*C3* (Park *et al.* (2009)), *LOX1* (Wang *et al.* (2011)) and *CD91* (Chalmers *et al.* (2010))).

The majority of the cohort after the year 2000 had high-resolution tissue typing for HLA Class I A,B,C and Class II DP,DQ and DR.

Statistical analysis

The impact of selected non-HLA genetic factors on survival, non-relapse mortality (NRM) (Iacobelli (2013)), relapse and GvHD was assessed using Kaplan-Meier, Cox regression and competing risks (Marubini & Valsecchi (1996)) techniques.

For modelling overall survival, biallelic SNPs were considered under the additive, dominant and recessive modes of inheritance (Chien, *et al.* (2012)). For a specific SNP, the mode showing the strongest association with survival was entered into the final model. All available genetic variables were used as candidates for Cox regression model building, these were entered into a stepwise variable selection modelling procedure alongside the EBMT risk score. A significance level of 0.05 was used for variable entry and 0.1 for removal.

Predictive accuracy of the models was quantified using the concordance index (C-index) (Harrell *et al.* (1982)) and prediction error (i.e. 0.632+ bootstrap estimator) (Gerds & Schumacher (2007)). A larger value of the concordance index (C-index) means that lower risk score correlates with longer survival time (C-index=1 means perfect predictive discrimination; C-index=0.5 means no predictive discrimination). The method of prediction error curve generation utilises a bootstrapping procedure whereby a model is generated using training data and validated using a test data set. The procedure is repeated 1000 times and, for each repetition, a training set of size N is extracted from the full data set (also of size N) using sampling with replacement; the remaining data (not used in the training sample) act as a test set. For each iteration, differences are found between actual and predicted responses in the test set and essentially an average residual is found. When several models are compared, a lower prediction error curve signifies that the associated model has a better predictive performance

R (v3.1.2), Minitab (v17), SAS (v9.4) and SPSS-23 were used for the computations.

Results

Clinical Characteristics

As transplant protocols have changed with time, we assessed clinical differences in patients treated up to and after the year 2000. The Chi-square test was employed for this task utilising exact *P*-values (Table 1).

The majority of patients in this study underwent HSCT after the year 2000 (63%). Transplants post year 2000 involved more patients and donors over 40 years of age, more HLA-matched unrelated donors, more peripheral blood stem cell transplants, more T cell depleted transplants and more reduced intensity conditioning (RIC) protocols (Table 1).

Analysis of Genetic Factors

Overall survival was initially studied. Kaplan-Meier survival curves and associated log rank tests were generated for each individual gene polymorphism. Variables showing a significant difference between survival curves (*P*-value < 0.05) are displayed in Table 2 and include *MAL* (rs8177374), the glucocorticoid receptor (GCR) haplotype (consisting of rs6198, rs33389 and rs33388), *IL1RN* (rs419598) and *IL4* (rs2243250) in the patient together with the *IL10* haplotype (consisting of rs1800896, rs1800871 and rs1800872) and *MDR1* (rs1045642) in the donor.

Model Construction

The multivariate Cox regression model was generated with N=204 (a reduction from N=458 due to missing data). A missing value assessment was carried out to check that there was no difference between the group of cases omitted and the group of cases included for statistical analysis. The difference between survival functions for the two groups was non significant (*P*-value = 0.376) (results not shown). Additional results showing the comparison of the clinical data are displayed in Supplementary Table 3.

Single nucleotide polymorphisms (SNPs) in three genes were found to improve the goodness of fit of the model when viewed alongside the EBMT risk score. Presence of *MAL* (rs8177374) allele T in the patient (HR:1.54, 95% CI:1.04-2.30), absence of the GCR haplotype (consisting of rs6198, rs33389 and rs33388) ACT in the patient (HR:0.67, 95% CI:0.46-0.98) and absence of heat shock protein 70-hom (HSP70-hom +2437) (rs2227956) allele C in the patient (HR:0.56, 95% CI:0.36-0.88) were associated with decreased survival time (Table 3).

Univariate log rank tests for the individual predictors and the associated Kaplan-Meier plots are shown in Supplementary Figure 1. Every predictor showed a log rank statistic with *P*-value <0.05 except patient HSP70-hom (+2437) (rs2227956) allele C (*P*-value = 0.245). This variable was therefore more weakly associated with overall survival when viewed on its own but an important predictor when viewed alongside other variables. Predictors having univariate

statistics with *P*-value < 0.25 are deemed to be satisfactory candidate variables in multivariate modelling (Hosmer & Lemeshow (2000)).

The resulting model enabled an integer score (clinical-genetic-score) to be assigned to each patient. The clinical-genetic-score was calculated via a summation process of the model's individual elements (Supplementary Section A). To make the scoring instrument easier to use in a clinical setting, individual elements were obtained by dividing the regression coefficients by the coefficient of the EBMT risk score and rounding to the nearest whole number (Table 3, Supplementary Section A). A higher risk score indicated a worse prognosis.

For the cases in the study set, the EBMT risk score ranged from 0 to 7 (low to high risk); clinical-genetic-scores ranged from 1 to 13 (no patient had the minimum possible score of 0 or maximum possible score of 14). After observing resulting Kaplan-Meier curves for the individual clinical-genetic-scores and grouping categories which naturally fell together (results not shown), 'low risk' could be categorized as having scores 1-6, 'intermediate risk': scores 7-8 and 'high risk': scores 9-13. A Kaplan-Meier plot displaying good separation of the three resulting survival curves is shown in Figure 1A. Furthermore the hazard ratio (HR) in a Cox regression significantly increased when compared to the low risk group (intermediate risk HR: 1.70, 95% CI: 1.10-2.63; high risk HR: 3.25, 95% CI: 2.06-5.12). A Kaplan-Meier plot for the EBMT risk score is also provided for comparison (Figure 1B), it was clear that the risk categories of the clinical-genetic-score had better separated (and consistently ordered) survival curves when compared to those of the EBMT risk score for this data set.

Next, prediction error curves (utilising a bootstrapping procedure and incorporating validation sets) were used to compare a model containing the single EBMT risk score, a model containing EBMT risk score and the three polymorphisms and a model containing no factors (i.e. the Kaplan-Meier curve). A comparison of each model is given in Figure 2. The lower prediction error curve for the model containing EBMT risk score and the three

polymorphisms signified that it had a better predictive value. Furthermore, the C-index showed that the risk score groups generated from a model containing SNPs correlated better with actual survival time when compared with a model containing the EBMT risk score alone (0.62 versus 0.59, *P*-value=0.012), the hazard ratio was also higher (HR: 1.80, 95% CI:1.43-2.26 versus HR: 1.21, 95% CI: 1.10-1.40).

Association with Other Outcomes

After further analysis of the SNPs for association with NRM and relapse, Gray's test (Gray (1988)) indicated that presence of *MAL* (rs8177374) allele T in the patient (P-value = 0.031) and absence of the GCR haplotype (consisting of rs6198, rs33389 and rs33388) ACT in the patient (P-value = 0.026) were associated with an increased incidence of relapse (Supplementary Tables 4A-B). The genetic variables were also studied for their association with acute GvHD I-IV, II-IV, III-IV and chronic GvHD. It was found that the presence of the GCR haplotype (consisting of rs6198, rs33389 and rs33388) ACT in the patient was weakly associated with chronic GvHD (P-value = 0.075).

Discussion

One of the problems of SNP association studies in HSCT is the heterogeneity of the cohort in terms of diagnosis, type of conditioning, GvHD prophylaxis and donor type. For this reason, this study focused on one disease group, acute leukaemia. For this disease, a risk scoring instrument was established which included three polymorphisms together with the EBMT risk score. Presence of *MAL* (rs8177374) allele T in the patient, absence of the GCR haplotype (consisting of rs6198, rs33389 and rs33388) ACT in the patient and absence of HSP70-hom +2437 (rs2227956) allele C in the patient were associated with decreased survival. The subsequent score assigned to each patient was termed the clinical-genetic-score and could be easily derived by means of summation of risk score points assigned to each of the model's elements. Moreover, in practice, polymorphisms in the patient could be typed at

the same time as donor tissue is typed (pre-transplant) and this makes the clinical-genetic-score no less accessible than the EBMT risk score. Patients could be assigned to low, intermediate and high risk score categories based on their clinical-genetic-score and identification of patients in the high risk group would ensure that they had a reduced time to transplant from diagnosis, younger donors, fully matched HLA donors (Kollman *et al.* (2001)) (all associated with improved survival) and increased monitoring for GvHD/infection.

As expected, there was a strong correlation between the EBMT risk score and clinical-genetic-score (Spearman's r = 0.676; P-value < 0.0005, N=204) but validation of the models via a bootstrapping procedure indicated that the clinical-genetic-score had improved predictive value when compared to the EBMT risk score alone and, additionally, prognostic risk categories derived from the clinical-genetic-score showed more distinct separation of survival curves. This result also held when data were divided according to transplant date with the clinical-genetic-score categories having a higher C-index when compared to the EBMT risk score (0.62 versus 0.57, up to year 2000; 0.63 versus 0.61, post year 2000) and displaying consistent order as regards pattern of survival curves (Supplementary Figures 2A-D). Additionally, the clinical-genetic-score appeared to highlight how transplants post 2000 have resulted in improved probability of survival for the intermediate risk group (Supplementary Figure 2B).

Holtick *et al.* (2008) found that patients homozygous for the T allele of *MAL* (rs8177374) had an increased risk of relapse whilst, in the donors, the presence of the T allele resulted in less cGvHD. Interestingly, Rocha *et al.* (2007) found, that presence of the T allele of rs8177374 in donors resulted in less transplant related mortality. The T allele is regarded in the literature as the inflammatory allele and individuals heterozygous for this SNP have increased protection from infection. Patients transplanted from donors with the T allele have also been

shown to have a lower incidence of fungal infections, lower incidence of acute GvHD and improved overall survival (Rocha, *et al.* (2007)). The *MAL* protein was originally identified in intermediate and late stages of T-lymphocyte differentiation (Alonso & Weissman (1987)) and is an internal component of glycolipid-enriched membrane domains in T-lymphocytes (Millán *et al.* (1997)). It may be important in T cell signalling and thus associated with reduced or improved survival depending on the role of the T cells e.g. associated GvL or GvHD effects. It is also important in the innate immune response and it is involved in Toll-like receptor (*TLR*) 2 and 4 signalling (Yamamoto *et al.* (2002)).

The GCR binds glucocorticoids in the cytoplasm and transports them to the nucleus where they influence the expression of many inflammatory genes such as interleukin (*IL*) *IL1, IL2, IL3, IL6, IL8* and *TNF* causing their down regulation (Chikanza *et al.* (2003)). In addition, glucocorticoids can also induce the transcription of anti-inflammatory genes including *IL4, IL10* and *IL13* (Chikanza, *et al.* (2003)). The ACT haplotype is made up of two of the putative GC resistant, inflammatory alleles, the C allele of rs33389 and the T allele of rs33388.

There is a well-documented association between the occurrence of cGvHD and reduced relapse risk in HSCT (Signori *et al.* (2012)) and this is borne out in our results with the ACT haplotype of the GCR in the patient showing a trend to association with cGvHD and the absence of this haplotype also being associated with decreased survival and with increased relapse. This putative inflammatory haplotype might be aiding the graft versus leukaemia (GvL) effect as well as cGvHD and therefore the absence of this haplotype would result in a more immunosuppressed phenotype in the patient leading to decreased survival.

There was a trend towards an association between recipient HSP70-hom +2437 (rs2227956) genotype TT and incidence of acute GvHD (P-value = 0.127). Past studies have shown a relationship between another HSP70-hom SNP, rs2075800, in recipients and GvHD with

genotype AA presenting more frequently with grade II to IV toxic lesions and acute GvHD (Bogunia-Kubik & Lange (2005)). Another study has shown an association between HSP70-hom +2437 (rs2227956) and treatment related mortality with the TT genotype in patients displaying a protective effect but the study was limited to a small cohort (N=147) (Kim *et al.* (2010)). HSP70-hom belongs to a highly conserved family of proteins, intracellular HSPs function as molecular chaperones facilitating the folding or transport of other proteins following physical or chemical stress (Matouschek (2003)). Extracellular HSPs are important in the cellular immune response acting as carrier molecules for antigen peptides (Srivastava *et al.* (1994)) and they may by themselves activate the innate immune response (Asea *et al.* (2000)).

This AL data set forms part of a previously analysed larger cohort which included other disease types: CML, lymphoma, plasma cell neoplasia, myelodysplasia syndrome and chronic myelomonocytic leukaemia (Balavarca, *et al.* (2015)). Here, a prognostic index for overall survival was also derived using the EBMT risk score and additional genetic polymorphisms (haplo-genotype ACC/ACC of *IL10* in donors, *MAL* rs8177374 in patients, *ESR1* rs9340799 in patients and *IL6* rs1800795 in donors). *MAL* rs8177374 in patients is common to both models with allele T being associated with reduced overall survival. Additionally, univariate similarities between the two cohorts were also present with a protective effect being observed for *IL10* haplo-genotype ACC/ACC, *IL4* rs2243250 (allele T) and GCR rs33388 (allele T) and a reduced survival effect for *MAL* rs8177374 (allele T) and *IL10* rs1800896 (allele G). Although, the purpose of the current (acute leukaemia) study is not to directly contrast and compare the two statistical models, we did find clear similarities when they were viewed alongside the established EBMT risk score (better differentiation of risk groups, increased predictive value and clear identification of improved survival probability for patients transplanted after 2000). Both studies hence illustrate the

importance of including genetic as well as clinical information in a pre-transplant risk scoring procedure and reflect the influence of polymorphisms which suppress the immune response giving rise to reduced survival (possibly due to reduced GvL effects) (Dickinson *et al.* (2010), Pearce *et al.* (2012), Balavarca, et al. (2015)).

In conclusion, the established EBMT risk score, although easily accessible pre-transplant, is recognised as not being optimal. Utilising clinical and genetic factors could be one way to improve pre-transplant clinical risk assessment. In a recent review, the relationship of the EBMT risk score with outcome was indeed shown to depend on other factors such as CMV status and Karnofsky score (Gratwohl (2012)). In future, further SNPs could be identified via meta-analysis and these could be viewed in conjunction with specific protein levels and/or mRNA and microRNA levels of putative biomarkers. In addition, large cohort analysis could enable combined patient and donor SNPs to be used as supplementary predictors in HSCT.

Acknowledgements

The authors would like to thank Dr Clare Lendrem for database management.

References

Alonso, M. A. & Weissman, S. M. (1987) CDNA cloning and sequence of MAL, a hydrophobic protein associated with human T-cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 1997.

Asea, A., Kraeft, S. K., Kurt-Jones, E. A., Stevenson, M. A., Chen, L. B., Finberg, R. W. *et al.* (2000) HSP70 stimulates cytokine production through a CD14-dependant pathway, demonstrating its dual role as a chaperone and cytokine. *Nature Medicine*, **6**, 435.

Balavarca, Y., Pearce, K., Norden, J., Collin, M., Jackson, G., Holler, E., *et al.* (2015) Predicting survival using clinical risk scores and non-HLA immunogenetics. *Bone Marrow Transplantation*, **50**, 1445.

Bogunia-Kubik, K. & Lange, A. (2005) HSP70-hom gene polymorphism in allogeneic hematopoietic stem-cell transplant recipients correlates with the development of acute graft-versus-host disease. *Transplantation*, **79**, 815.

Branford, S. (2007) Chronic myeloid leukemia: Molecular monitoring in clinical practice. *Hematology/American Society Hematololgy Educuation Program*, 376.

Cargill, M., Schrodi, S. J., Chang, M., Garcia, V. E., Brandon, R., Callis, K. P., et al. (2007) A large-scale genetic association study confirms il12b and leads to the identification of il23r as psoriasis-risk genes. *American Journal of Human Genetics*, **80**, 273.

Cavet, J., Dickinson, A. M., Norden, J., Taylor, P. R., Jackson, G. H. & Middleton, P. G. (2001) Interferon-gamma and interleukin-6 gene polymorphisms associate with graft-versus-host disease in HLA-matched sibling bone marrow transplantation. *Blood*, **98**, 1594.

Chalmers, K. A., Barker, R., Passmore, P. A., Panza, F., Seripa, D., Solfrizzi, V. *et al.* (2010) LRP-1 variation is not associated with risk of alzheimer's disease. *International Journal of Molecular Epidemiology and Genetics*, **1**, 104.

Chien, J. W., Zhang, X. C., Fan, W., Wang, H., Zhao, L. P., Martin, P. J., *et al.* (2012) Evaluation of published single nucleotide polymorphisms associated with acute GvHD. *Blood*, **119**, 5311.

Chikanza, I. C., Kozaci, D. & Chernajovsky, Y. (2003) The molecular and cellular basis of corticosteroid resistance. *Journal of Endocrinology*, **179**, 301.

Copelan, E. A. (2006) Hematopoietic stem-cell transplantation. *New England Journal of Medicine*, **354**, 1813.

Derijk, R. H., Schaaf, M. J., Turner, G., Datson, N. A., Vreugdenhil, E., Cidlowski, J. *et al.* (2001) A human glucocorticoid receptor gene variant that increases the stability of the glucocorticoid receptor beta-isoform mRNA is associated with rheumatoid arthritis. *Journal of Rheumatology*, **28**, 2383.

Dickinson, A. M. (2008) Non-HLA genetics and predicting outcome in HSCT. *International Journal of Immunogenetics*, **35**, 375.

Dickinson, A. M., Pearce, K. F., Norden, J., O'Brien, S. G., Holler, E., Bickeboller, H. *et al.* (2010) Impact of genomic risk factors on outcome after hematopoietic stem cell transplantation for patients with chronic myeloid leukemia. *Haematologica*, **95**, 922.

Feinstein, L., Sandmaier, B., Maloney, D., McSweeney, P. A., Maris, M., Flowers, C. *et al.* (2001) Nonmyeloablative hematopoietic cell transplantation. Replacing high-dose cytotoxic therapy by the graft-versus-tumor effect. *Annals of New York Academy of Sciences*, **938**, 328.

Gerds, T. A. & Schumacher, M. (2007) Efron-type measures of prediction error for survival analysis. *Biometrics*, **63**, 1283.

Gratwohl, A. (2012) The EBMT risk score. *Bone Marrow Transplantation*, **47**, 749.

Gratwohl, A., Baldomero, H., Frauendorfer, K., Urbano-Ispizua, A. & Niederwieser, D. (2007) Results of the EBMT activity survey 2005 on haematopoietic stem cell transplantation: Focus on increasing use of unrelated donors. *Bone Marrow Transplantation*, **39**, 71.

Gratwohl, A., Stern, M., Brand, R., Apperley, J., Baldomero, H., de Witte, T. *et al.* (2009) Risk score for outcome after allogeneic hematopoietic stem cell transplantation: A retrospective analysis. *Cancer*, **115**, 4715.

Graves, P. E., Kabesch, M., Halonen, M., Holberg, C. J., Baldini, M., Fritzsch, C. *et al.* (2000) A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children. *Journal of Allergy and Clinical Immunology*, **105**, 506.

Gray , R. (1988) A class of K sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, **16**, 1141.

Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. (1982) Evaluating the yield of medical tests. *The Journal of the American Medical Association*, **247**, 2543.

Holler, E., Rogler, G., Herfarth, H., Brenmoehl, J., Wild, P. J., Hahn, J. *et al.* (2004) Both donor and recipient NOD2/CARD15 mutations associate with transplant-related mortality and GvHD following allogeneic stem cell transplantation. *Blood*, **104**, 889.

Hosmer, D. W. & Lemeshow, S. (2000) *Applied logistic regression*, 2nd edn. John Wiley and Sons, Chichester, England.

Hurme, M. & Santtila, S. (1998) IL-1 receptor antagonist (IL-1Ra) plasma levels are co-ordinately regulated by both IL-1Ra and IL-1 beta genes. *European Journal of Immunology*, **28**, 2598.

Iacobelli, S. (2013) Suggestions on the use of statistical methodologies in studies of the European group for blood and marrow transplantation. *Bone Marrow Transplantation*, **48**, S1.

Imboden, M., Nieters, A., Bircher, A. J., Brutsche, M., Becker, N., Wjst, M. *et al.* (2006) Cytokine gene polymorphisms and atopic disease in two European cohorts. (ECRHS-Basel and SAPALDIA). *Clinial and Molecular Allergy*, **4**, 9.

Kim, D. H., Park, J. Y., Sohn, S. K., Lee, N. Y., Suh, J. S. & Lee, K. B. (2006) The association between multidrug resistance-1 gene polymorphisms and outcomes of allogeneic HLA-identical stem cell transplantation. *Haematologica*, **91**, 848.

Kim, I., Kim, J. H., Rhee, J. Y., Kim, J. W., Cho, H. J., Cho, E. Y. *et al.* (2010) Patient HSP70-hom TG haplotype is associated with decreased transplant-related mortality and improved survival after sibling HLA-matched hematopoietic stem cell transplantation. *Clinical Transplantation*, **24**, 459.

Klein, W., Tromm, A., Griga, T., Fricke, H., Folwaczny, C., Hocke, M. *et al.* (2002) A polymorphism in the CD14 gene is associated with Crohn disease. *Scandinavian Journal of Gastroenterology*, **37**, 189.

Kollman, C., Howe, C. W., Anasetti, C., Antin, J. H., Davies, S. M., Filipovich, A. H. *et al.* (2001) Donor characteristics as risk factors in recipients after transplantation of bone marrow from unrelated donors: The effect of donor age. *Blood*, **98**, 2043.

Marubini, E. & Valsecchi, M. G. (1996) Analysing survival data from clinical trials and observational studies. John Wiley & Sons, Chichester, England.

Matouschek, A. (2003) Protein unfolding-an important process in vivo? *Current Opinion in Structural Biology*, **13**, 98.

Middleton, P. G., Cullup, H., Dickinson, A. M., Norden, J., Jackson, G. H., Taylor, P. R. *et al.* (2002) Vitamin D receptor gene polymorphism associates with graft-versus-host disease and survival in HLA-matched sibling allogeneic bone marrow transplantation. *Bone Marrow Transplantation*, **30**, 223.

Middleton, P. G., Norden, J., Cullup, H., Cavet, J., Jackson, G. H., Taylor, P. R. *et al.* (2003) Oestrogen receptor alpha gene polymorphism associates with occurrence of graft-versus-host disease and reduced survival in HLA-matched sib-allo BMT. *Bone Marrow Transplantation*, **32**, 41.

Millán, J., Puertollano, R., Fan, L., Rancaño, C. & Alonso, M. A. (1997) The MAL proteolipid is a component of the detergent-insoluble membrane subdomains of human T-lymphocytes. *Biochemical Journal*, **321**, 247.

Morse, H. R., Olomolaiye, O. O., Wood, N. A., Keen, L. J. & Bidwell, J. L. (1999) Induced heteroduplex genotyping of TNF-alpha, IL-1 Beta, IL-6 and IL-10 polymorphisms associated with transcriptional regulation. *Cytokine*, **11**, 789.

Mullally, A. & Ritz, J. (2007) Beyond HLA: The significance of genomic variation for allogeneic hematopoietic stem cell transplantation. *Blood*, **109**, 1355.

Novota, P., Zinöcker, S., Norden, J., Wang, X. N., Sviland, L., Opitz, L. *et al.* (2011) Expression profiling of major histocompatibility and natural killer complex genes reveals candidates for controlling risk of graft versus host disease. *PLoS ONE*, **6**, e16582.

Park, K. H., Fridley, B. L., Ryu, E., Tosakulwong, N. & Edwards, A. O. (2009) Complement component 3 (C3) haplotypes and risk of advanced age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, **50**, 3386.

Pearce, K. F., Lee, S. J., Haagenson, M., Petersdorf, E. W., Norden, J., Collin, M. P. *et al.* (2012) Analysis of non-HLA genomic risk factors in HLA-matched unrelated donor hematopoietic cell transplantation for chronic myeloid leukemia. *Haematologica*, **97**, 1014.

Pravica, V., Asderakis, A., Perrey, C., Hajeer, A., Sinnott, P. J. & Hutchinson, I. V. (1999) In vitro production of IFN-gamma correlates with CA repeat polymorphism in the human IFN-gamma gene. *European Journal of Immunogenetics*, **26**, 1.

Rocha, V., Porcher, R., Kabbara, N., de Latour, R. P., Robin, M., Chaves, W. *et al.* (2007) Mutant TIRAP gene polymorphism is associated with outcomes in HLA genoidentical bone marrow transplants recipients with leukemia. In: *Blood ASH Annual Meeting Abstracts.* Abstract 327.

Signori, A., Crocchiolo, R., Oneto, R., Sacchi, N., Sormani, M. P., Fagioli, F. *et al.* (2012) Chronic GvHD is associated with inferior relapse risk irrespective of stem cell source among patients receiving transplantation from unrelated donors. *Bone Marrow Transplantation*, **47**, 1474.

Srivastava, P. K., Udono, H., Blachere, N. E. & Li, Z. (1994) Heat shock proteins transfer peptides during antigen processing and CTL priming. *Immunogenetics*, **39**, 93.

Stark, G. L., Dickinson, A. M., Jackson, G. H., Taylor, P. R., Proctor, S. J. & Middleton, P. G. (2003) Tumour necrosis factor receptor type II 196M/R genotype correlates with circulating soluble receptor levels in normal subjects and with graft-versus-host disease after sibling allogeneic bone marrow transplantation. *Transplantation*, **76**, 1742.

Stevens, A., Ray, D. W., Zeggini, E., John, S., Richards, H. L., Griffiths, C. E. *et al.* (2004) Glucocorticoid sensitivity is determined by a specific glucocorticoid receptor haplotype. *Journal of Clinical Endocrinology and Metabolism*, **89**, 892.

Walley, A. J. & Cookson, W. O. (1996) Investigation of an interleukin-4 promoter polymorphism for associations with asthma and atopy. *Journal of Medical Genetics*, **33**, 689.

Wang, L., Yanuck, D., Beecham, A., Gardener, H., Slifer, S., Blanton, S. H. *et al.* (2011) A candidate gene study revealed sex-specific association between the OLR1 gene and carotid plaque. *Stroke*, **42**, 588.

Yamamoto, M., Sato, S., Hemmi, H., Sanjo, H., Uematsu, S., Kaisho, T. *et al.* (2002) Essential role for TIRAP in activation of the signalling cascade shared by TLR2 and TLR4. *Nature*, **420**, 324.

Yee, L. J., Im, K., Borg, B., Yang, H. & Liang, T. J. (2009) Interleukin-6 (IL-6) haplotypes and the response to therapy of chronic hepatitis c virus infection. *Genes and immunity*, **10**, 365.

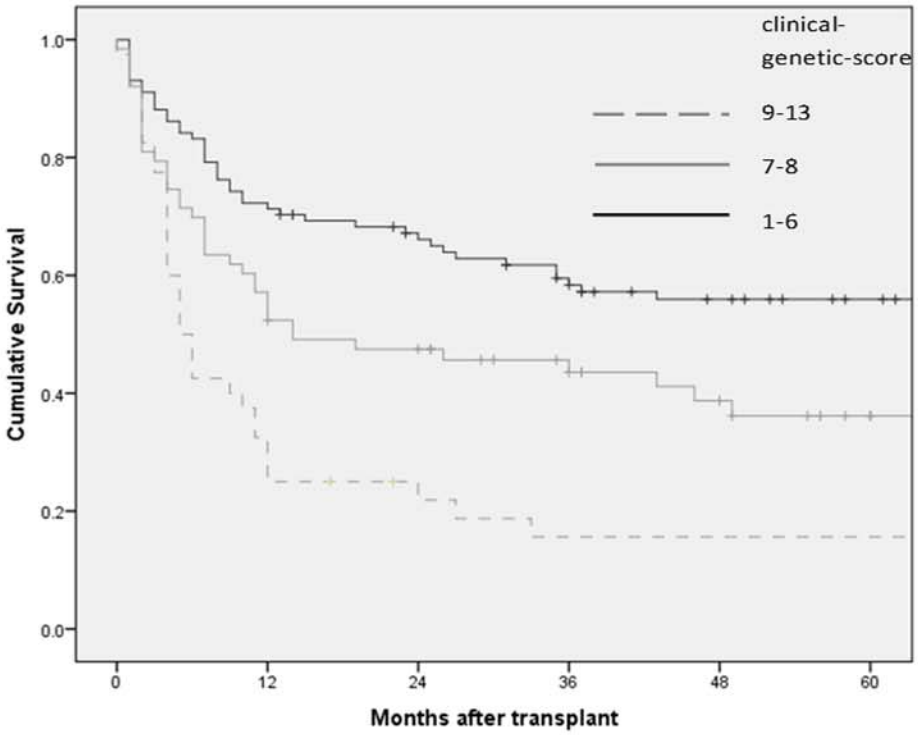
Figure Legends

Figure 1 Kaplan-Meier survival plots for risk groups. Figure A Derived from the clinical-genetic-score. Grey dashed line = scores 9-13, grey solid line = scores 7-8, black solid line = scores 1-6. Figure B Derived from the EBMT risk score. Grey dotted line = scores 6-7, grey dashed line = score 5, grey solid line = score 4, black dotted line = score 3, black dashed line = score 2, black solid line = score 0-1. Crosses represent censored observations. N=204. It is illustrated that the risk categories of the clinical-genetic-score had better separated (and consistently ordered) survival curves when compared to those of the EBMT risk score.

Figure 2 Plot of prediction error curve for model including EBMT risk score and 3 polymorphisms compared to (i) EBMT risk score and (ii) a model containing no factors (Kaplan-Meier curve). N=204. Solid black line: Kaplan-Meier curve; solid grey line: EBMT risk score; dashed grey line: model including EBMT risk score and 3 polymorphisms. The lower prediction error curve for the model containing EBMT risk score and the three polymorphisms showed that it had a better predictive value.

Figure 1

A


clinical-genetic-score
9-13
7-8
1-6
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
60
Months after transplant

B

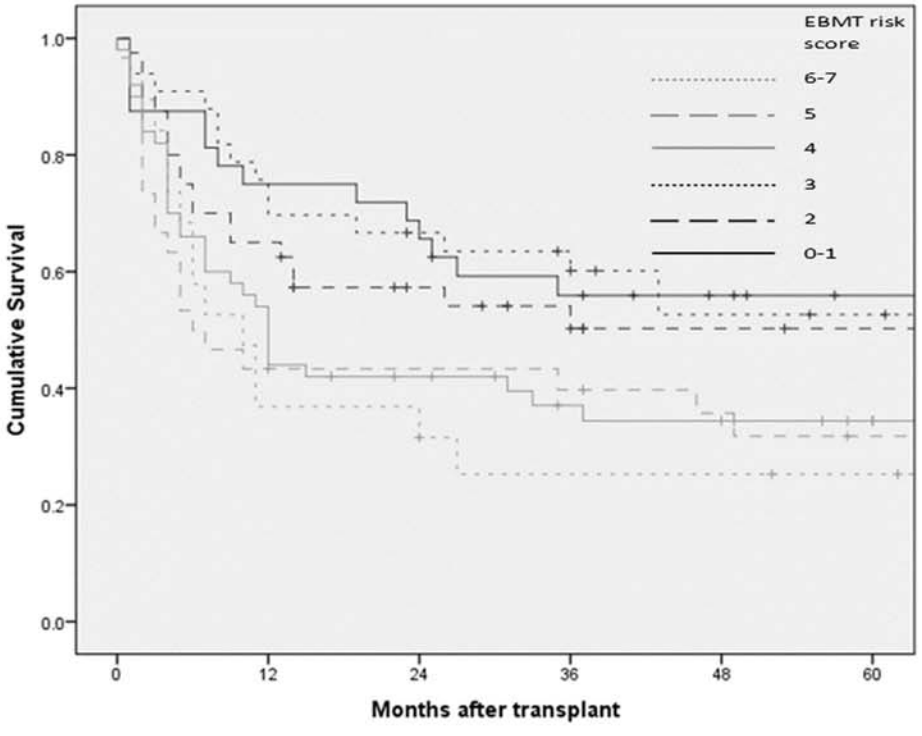

EBMT risk score
6-7
5
4
3
2
0-1
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
60
Months after transplant

Figure 2

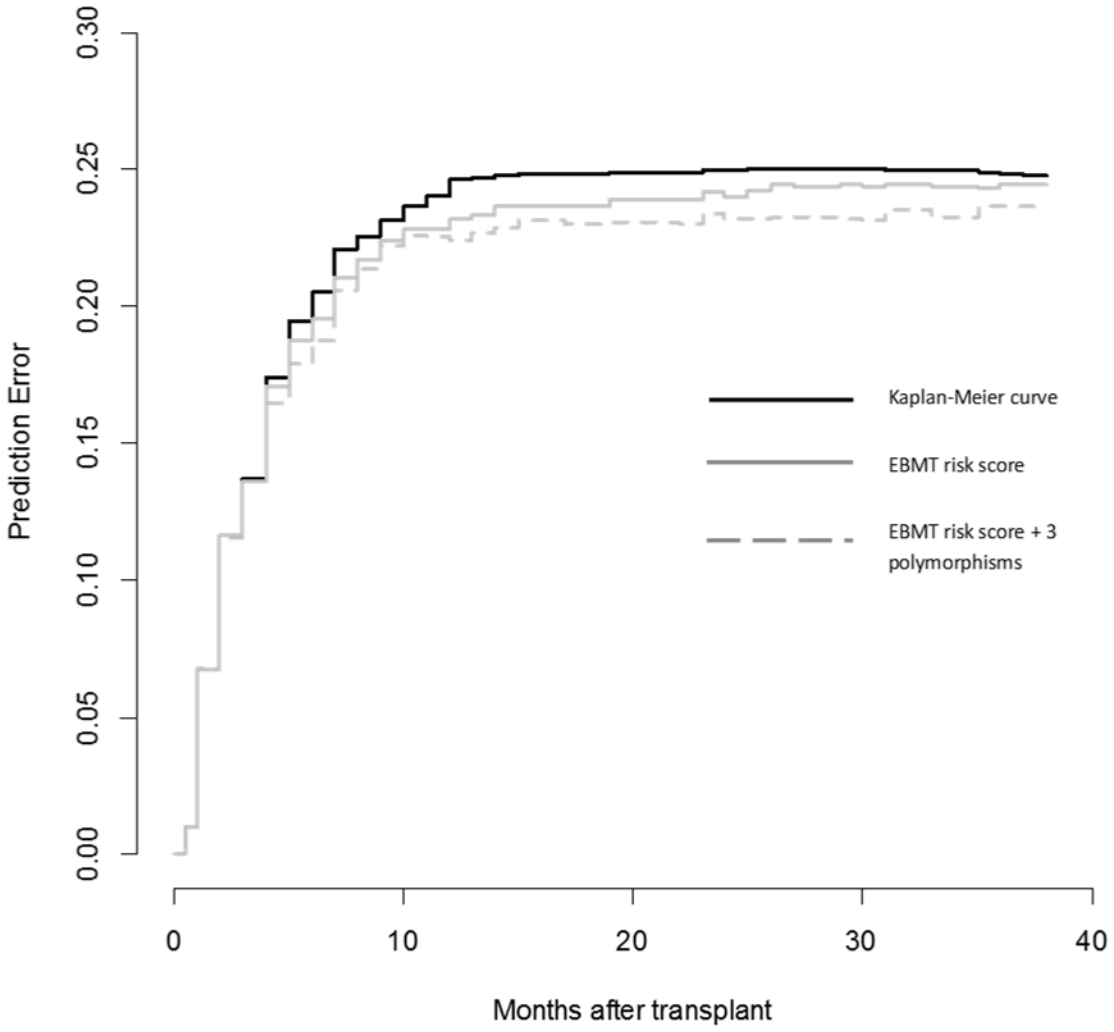

Prediction Error
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
Months after transplant
Kaplan-Meier curve
EBMT risk score
EBMT risk score + 3 polymorphisms

Table 1 Clinical characteristics of patients and donors prior to HSCT (N=458)

Factors	Categories	No.	(%)	Year of Transplant ≤ 2000 No.	(%)	Year of Transplant >2000 No.	(%)	*P*-value[c]
No of patient/donor				171		287		
Age of patient at transplantation (years)[a,b]	<20 years	26	(5.7)	17	(9.9)	9	(3.1)	<0.001
	20-40 years	226	(49.3)	107	(62.6)	119	(41.5)	
	>40 years	206	(45.0)	47	(27.5)	159	(55.4)	
Age of donor (years)[b]	<20 years	26	(5.7)	21	(12.6)	5	(1.8)	<0.001
	20-40 years	261	(57.0)	97	(58.1)	164	(59.0)	
	>40 years	158	(34.5)	49	(29.3)	109	(39.2)	
	NA	13	(2.8)	4		9		
Patient/donor gender combination[a]	Male/female	93	(79.5)	33	(19.3)	60	(21.0)	0.719
	Other	364	(20.3)	138	(80.7)	226	(79.0)	
	NA	1	(0.2)	0		1		
Type of donor[a,b]	Sibling	241	(47.4)	113	(66.1)	128	(44.6)	<0.001
	HLA-matched unrelated(MUD)	217	(52.6)	58	(33.9)	159	(55.4)	
Haematological disease	Acute Biphenotypic leukaemia	3	(0.7)	2	(1.2)	1	(0.3)	0.547
	Acute Lymphoblastic leukaemia	123	(26.9)	47	(27.5)	76	(26.5)	
	Acute Myeloid leukaemia	330	(72.1)	122	(71.3)	208	(72.5)	
	Acute Undifferentiated leukaemia	2	(0.4)	0	(0.0)	2	(0.7)	
Source of stem cells[b]	Bone Marrow	207	(45.2)	138	(83.1)	69	(25.5)	<0.001
	Peripheral blood	229	(50.0)	28	(16.9)	201	(74.2)	
	Both sources	1	(0.2)	0	(0.0)	1	(0.4)	
	NA	21	(4.6)	5		16		
Patient/donor CMV status	Negative/Negative	136	(29.7)	56	(34.4)	80	(28.5)	0.202
	Other	308	(67.2)	107	(65.6)	201	(71.5)	
	NA	14	(3.1)	8		6		
Stage of disease at transplantation[a]	Early	185	(40.4)	77	(48.1)	108	(43.9)	0.715
	Intermediate	92	(20.1)	35	(21.9)	57	(23.2)	
	Late	129	(28.2)	48	(30.0)	81	(32.9)	
	NA	52	(11.4)	11		41		
Time from diagnosis to transplant[a]	≤12 months	330	(72.1)	124	(73.8)	206	(76.3)	0.570
	>12 months	108	(23.6)	44	(26.2)	64	(23.7)	
	NA	20	(4.4)	3		17		
T cell depletion[b]	T cell depletion	154	(33.6)	28	(16.4)	126	(43.9)	<0.001
	No T cell depletion	304	(66.4)	143	(83.6)	161	(56.1)	
Conditioning regimen[b]	Standard myeloablative	327	(71.4)	154	(90.1)	173	(60.3)	<0.001
	Reduced Intensity (RIC)	131	(28.6)	17	(9.9)	114	(39.7)	
Transplantation center[b]	Vienna/Prague[d]	136	(30.0)	44	(25.7)	92	(32.1)	<0.001
	Regensburg/Munich[d]	121	(26.4)	32	(18.7)	89	(31.0)	
	Newcastle	123	(26.9)	83	(48.5)	40	(13.9)	
	Rostock	31	(6.8)	9	(5.3)	22	(7.7)	
	Paris	28	(6.1)	0	(0.0)	28	(9.8)	
	Barcelona	19	(4.1)	3	(1.8)	16	(5.6)	

Abbreviations: HSCT = haematopoietic stem cell transplantation; NA = not available; No. = number.
[a]Clinical EBMT factor. Categories displayed are those which are necessary for the derivation of the EBMT risk score.
[b]Significant difference between time periods (*P*-value ≤ 0.05).

[c]*P*-values for Chi square (exact) test.
[d]The centers of Vienna and Prague worked in close collaboration with each other - hence, both centres used similar treatment procedures. Likewise, there were collaborations between the centres of Regensburg and Munich.

Table 2 Gene polymorphisms associated with overall survival with log rank test having *P*-value <0.05

Genes	Polymorphism[a]	Genotypes associated with decreased survival	*P*-value	MAF(%)[b]
P-*MAL*	rs8177374(T)	TT or TC	0.003	15
D-*IL10*	haplotype GCC	haplo-genotypes including GCC[c]	0.008	49
P-GCR	haplotype ACT	haplo-genotypes excluding ACT[d]	0.013	46
D-*MDR1*	rs1045642(C)	CC	0.026	44
P-*IL1RN*	rs419598(C)	CC or TC	0.030	25
D-*IL10*	haplo-genotype (ACC/ACC)	haplo-genotypes excluding (ACC/ACC)	0.038	26
P-*IL4*	rs2243250(T)	CC	0.043	16

Abbreviations: Patients (P-) = gene from patient; D- = gene from donor; MAF= Minor Allele Frequency.
[a]Polymorphisms ordered from lowest to highest P-value for log rank test. SNP rs numbers are followed by the minor allele.
[b]Minor allele or haplotype frequency of the respective polymorphism tested for association with overall survival.
[c]SNPS in the promoter region of the *IL10* gene give the haplotypes GCC, ACC and ATA. The SNP order for the haplotype designation is rs1800896, then rs1800871, then rs1800872. The result for haplotype GCC is therefore equivalent to that for the presence of the G allele in rs1800896.
[d]SNPs in the GCR gene give distinct haplotypes GCA,ACA,ATA and ACT. The SNP order for the haplotype designation is rs6198, then rs33389, then rs33388. The result for haplotype ACT is therefore equivalent to that for the presence of the T allele in rs33388.

Table 3 Cox regression model with EBMT risk score and multiple polymorphisms for overall survival (N=204)

Factors[a]	Coefficient[b]	*P*-value	Hazard Ratio[c]	Confidence Interval (95%)	Risk Score Element[d]
EBMT risk score	0.20	0.001	1.22	1.08-1.37	0-7[e]
P-*MAL* rs8177374(T) dom.	0.43	0.033	1.54[f]	1.04-2.30	TC or TT=2; CC=0
P-GCR hapolotype (rs6198, rs33389, rs33388) ACT	-0.39	0.049	0.67[f]	0.46-0.98	(ACT/GCA), (ACT/ACA),(ACT/ACT) or (ACT/ATA)=0; others=2
P-HSP70-hom(+2437) rs2227956(C) dom.	-0.57	0.012	0.56[f]	0.36-0.88	CC or TC=0; TT=3

Abbreviations: Patients (P-) = gene from patient.

[a] For the SNPs listed, the rs number is followed by the minor allele and the respective genetic model: additive (add.), dominant (dom.), or recessive (rec.).

[b] Regression coefficients from the Cox model, equivalent to the log of the hazard ratio.

[c] As EBMT risk score is evaluated using "histocompatibility" as one of its component elements and the percentage of related transplants receiving T cell depletion is lower than the percentage of unrelated transplants receiving T cell depletion, it may be argued that T cell depletion is exerting a confounding effect. However, after adjusting for T cell depletion in the model, Hazard Ratios for the model factors are virtually unchanged (1.18, 1.62, 0.68, 0.57 respectively).

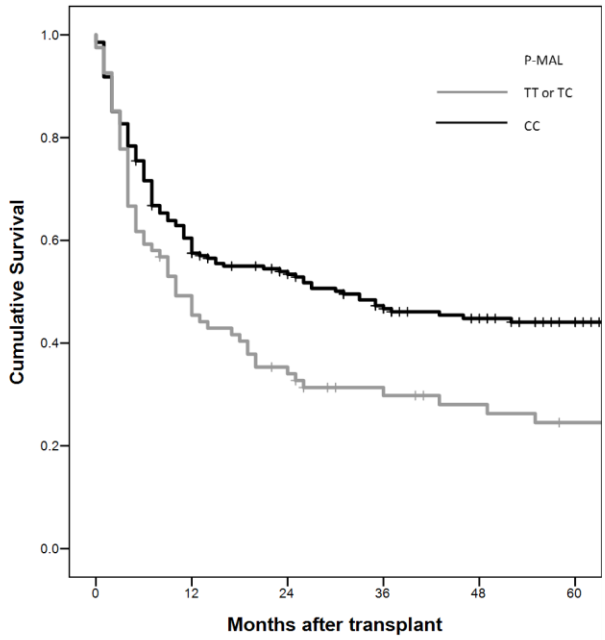
[d] Elements are summed to obtain risk score for each patient (Supplementary Section A). Minimum possible score = 0, maximum possible score=14.

[e] Range of score values for EBMT risk score (Gratwohl *et al.*, 2009).

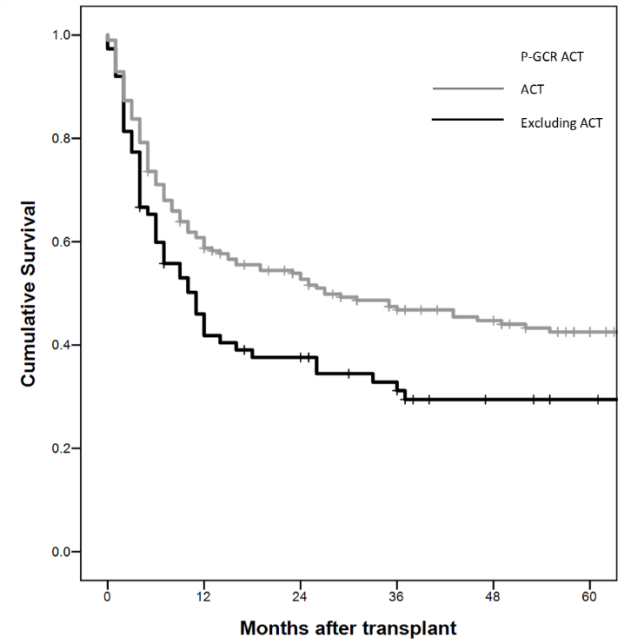
[f] Note the effect size, either protective or at risk, is larger than that of EBMT risk score.

Supplementary Figure 1 Association with overall survival of individual gene polymorphisms included in prognostic model (Kaplan-Meier plots). Figure A Patient *MAL* rs8177374: grey line represents TT or TC, black line represents CC; log rank test p=0.003 , Figure B Patient GCR hapolotype (rs6198, rs33388, rs33389) ACT: grey line represents haplo-genotypes including ACT, black line represents haplo-genotypes excluding ACT; log rank test p=0.013, Figure C Patient HSP70-hom(+2437) rs2227956(C): grey line represents CC or TC, black line represents TT; log rank test p=0.245.

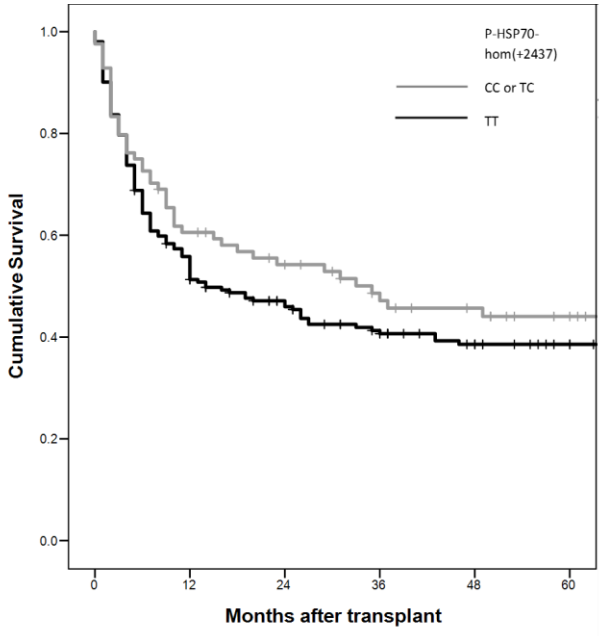
A



B



C

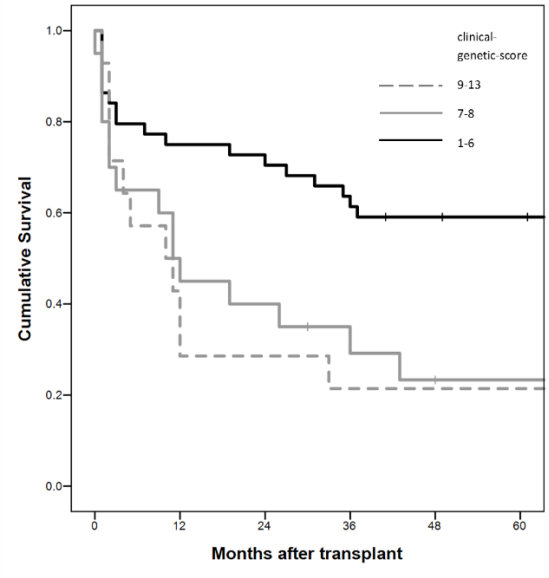

P-HSP70-hom(+2437)
CC or TC
TT
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
60
Months after transplant

Supplementary Figure 2

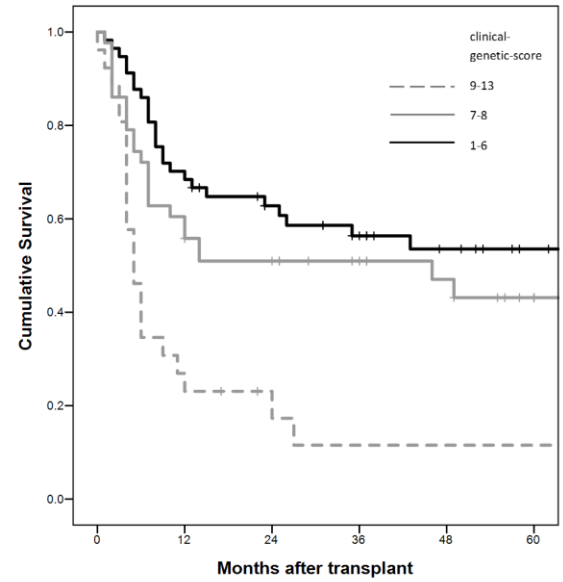
Kaplan-Meier survival plots for risk groups derived from the clinical-genetic-score: Figure 2A ≤ year 2000 (N=78), Figure 2B > year 2000 (N=126). Grey dashed line = scores 9-13, grey solid line = scores 7-8, black solid line = scores 1-6. Kaplan-Meier survival plots for risk groups derived from the EBMT risk score: Figure 2C ≤ year 2000 (N=78), Figure 2D > year 2000 (N=126). Grey dotted line = scores 6-7, grey dashed line = score 5, grey solid line = score 4, black dotted line = score 3, black dashed line = score 2, black solid line = score 0-1. Crosses represent censored observations.

It is illustrated that the risk categories of the clinical-genetic-score had consistently ordered survival curves when compared to those of the EBMT risk score.

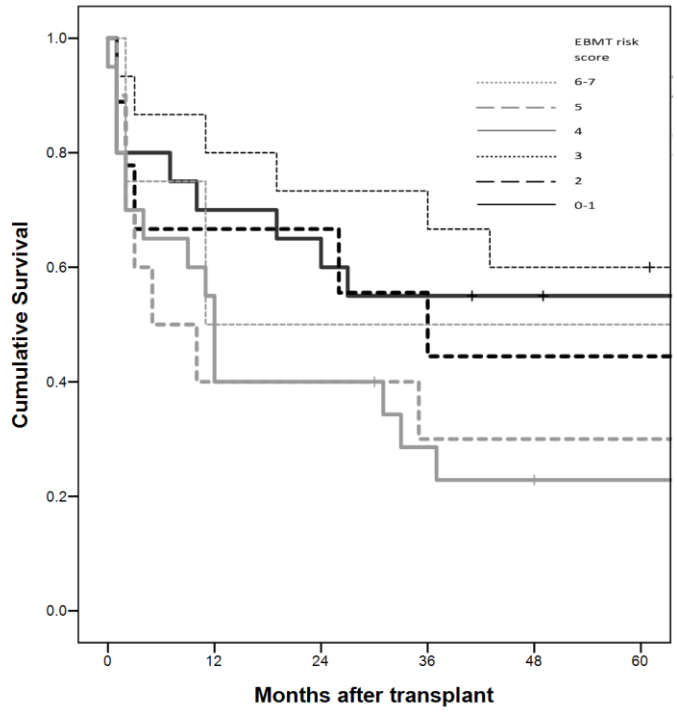
A



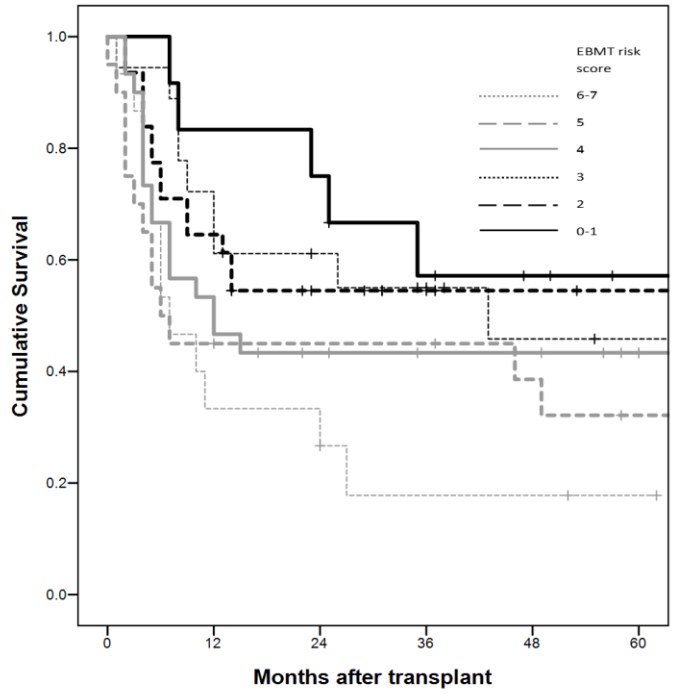
B



C


EBMT risk score
6-7
5
4
3
2
0-1
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
60
Months after transplant

D


EBMT risk score
6-7
5
4
3
2
0-1
Cumulative Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
12
24
36
48
60
Months after transplant

Supplementary Table 1 Categorisations of the individual clinical elements that are utilised in the derivation of the EBMT risk score.

Risk Factor		Score Points[b]
Age of Patient (in years)		
	<20	0
	20-40	1
	>40	2
Disease Stage		
	Early	0
	Intermediate	1
	Late	2
Time interval for diagnosis to transplant (in months)[a]		
	≤12	0
	>12	1
Donor Type		
	HLA-identical sibling donor	0
	Unrelated donor, other	1
Patient/donor gender combination[c]		
	Male patient/female donor	1
	All other	0

[a]does not apply for patient transplanted in first complete remission (score 0)
[b]Example of derivation: Patient < 20 years , intermediate disease stage, HSCT >12 months after diagnosis, unrelated donor, male patient female donor : EBMT risk score = 0 + 1 + 1 + 1 + 1 = 4

Supplementary Table 2 Description of non-HLA SNPs and other polymorphisms[a] (N=458).

Gene name	Chr.	SNP rs number	Allele pair	MA	MAF %		Sample Size[b]	
					patient	donor	patient	donor
Cluster determinant 14 (*CD14*)	5	rs2569190	G/A	A	49	46	295	308
Cluster determinant 91 (*LPR1*, *CD91*)	12	rs1799986	C/T	T	15	15	297	315
Complement component 3 (*C3*)	19	rs2230199	C/G	G	20	20	309	322
Estrogen receptor 1 (*ESR1*)	6	rs2234693	C/T	C	45	45	382	378
		rs9340799	G/A	G	40	37	378	372
Glucocorticoid receptor (*NR3C1*, GCR)	5	rs33389	C/T	T	16	16	298	329
		rs33388	T/A	T	45	46	303	322
		rs6198	G/A	G	15	17	303	321
Heat shock (*HSPA1L*, HSP70-hom)	6	rs2075800	G/A	A	35	35	248	234
protein 70 Hom		rs2227956	T/C	C	16	17	292	286
IL1 receptor antagonist (*IL1RN*)	2	rs419598	T/C	C	25	22	386	388
Interleukin 4 (*IL4*)	5	rs2243250	T/C	T	16	16	361	358
Interleukin 6 (*IL6*)	7	rs1800797	G/A	A	38	41	327	354
		rs1800796	C/G	C	6	6	348	370
		rs1800795	G/C	C	39	41	390	389
Interleukin 10 (*IL10*)	1	rs1800896	G/A	G	48	49	346	356
		rs1800872	A/C	A	25	25	362	367
Interleukin 12B (*IL12B*)	5	rs3212227	A/C	C	23	19	264	285
Interleukin 13 (*IL13*)	5	rs1800925	C/T	T	19	17	314	340
		rs20541	A/G	A	19	22	340	350
		rs1881457	C/A	C	18	18	339	359
Oxidized Low-Density (*OLR1*, *LOX1*)	12	rs11053646	G/C	G	7	8	300	324
Lipoprotein Receptor 1								
MyD88-adapter-like (*TIRAP*, *MAL*)	11	rs8177374	T/C	T	15	15	295	318
Multi drug resistance (*ABCB1*, *MDR1*)	7	rs1045642	C/T	C	48	44	280	292
Nucleotide-binding (*NOD2*)	16	rs2066844	C/T	T	5	6	322	328
oligomerization domain		rs2066845	C/G	C	2	1	322	328
containing 2		rs2066847	-/ C	C	2	3	322	328
Tumour necrosis factor (*TNF*)	6	rs1800629	G/A	A	15	16	346	343
Tumour necrosis factor (*TNFRSF1B*) receptor 2	1	rs1061622	T/G	G	25	24	355	373
Vitamin D receptor (*VDR*)	12	rs731236	T/C	C	39	39	370	387
		rs7975232	C/A	C	49	47	377	396

Abbreviations: Chr. = Chromosome number. MA = Minor allele. MAF = Minor allele frequency.
[a]Other non-HLA polymorphisms were from genes: *IFNG*, *MDR1*(three allelic), GCR(haplotype) and *IL10*(haplotype).
[b]Available sample size for patient's or donor's SNP.

Supplementary Table 3 Comparing clinical variables for the subgroup used for modelling with the subgroup not used for modelling. It was observed that there was no statistical difference between the group of cases omitted and the group of cases included for statistical analysis.

	Used for Modelling	Not Used for Modelling	*P*-value
	Mean (SD)	Mean (SD)	
Age of patient at transplantation (years)	38.7 (11.44)	39.1 (12.71)	0.709[1]
Age of donor at transplantation (years)	36.0 (10.53)	38.1 (12.02)	0.051[1]
Time from diagnosis to transplant (months)	14.1 (24.98)	10.3 (12.03)	0.051[1]
	%	%	
Female donor to male patient	19.6	20.9	0.815[2]
CMV positivity in either patient or donor	68.3	70.2	0.680[2]
Source of stem cells (Bone Marrow)	52.7	43.4	0.055[2]
HLA-matched unrelated (MUD) transplant	54.4	45.3	0.06[2]
Reduced Intensity Conditioning (RIC)	29.4	28.0	0.756[2]
T-cell depletion	33.8	33.5	1.000[2]
Late stage disease at transplantation	29.0	35.0	0.241[2]

Abbreviation: SD = standard deviation
[1] Two sample t-test (2 sided test)
[2] Fisher's Exact Test (2 sided test)

Supplementary Table 4A Cumulative incidence of NRM at 2 years and 5 years post transplant for each level of a SNP. Table 4B Cumulative incidence of relapse at 2 years and 5 years post transplant for each level of a SNP. Associated *P*-values for Gray's test are shown to establish if there is a significant difference between cumulative incidence curves. The presence of *MAL* (rs8177374) allele T in the patient and absence of the GCR haplotype ACT in the patient were associated with an increased incidence of relapse.

A

Genes[a]	Genotype/Haplotype	2 years CI[b]	5 years CI[b]	*P*-value[c]
P-*MAL* rs8177374(T)	CC	0.26	0.27	0.263
	TT, TC	0.28	0.36	
P-GCR hapolotype (rs6198, rs33388, rs33389) ACT	others	0.28	0.28	0.942
	(ACT/GCA), (ACT/ACA),(ACT/ACT) or (ACT/ATA)	0.26	0.31	
P-HSP70-hom(+2437) rs2227956(C)	TT	0.28	0.29	0.862
	CC,TC	0.27	0.33	

B

Genes[a]	Genotype/Haplotype	2 years CI[b]	5 years CI[b]	*P*-value[c]
P-*MAL* rs8177374(T)	CC	0.26	0.29	0.031
	TT, TC	0.39	0.39	
P-GCR hapolotype (rs6198, rs33388, rs33389) ACT	others	0.40	0.40	0.026
	(ACT/GCA), (ACT/ACA),(ACT/ACT) or (ACT/ATA)	0.25	0.27	
P-HSP70-hom(+2437) rs2227956(C)	TT	0.31	0.31	0.160
	CC,TC	0.20	0.23	

Abbreviations: Patients (P-) = gene from patient; D- = gene from donor.
[a] For the SNPs listed, the rs number is followed by the minor allele.
[b]CI = cumulative incidence.
[c]Gray's test.

Supplementary Section

A

Risk score elements to be used in the derivation of the clinical-genetic-score (final column, Table 2) are derived from the coefficients of the Cox regression model. The model took the form:

0.196 x EBMT risk score + 0.434 x *MAL* rs8177374 - 0.394 x GCR haplotype ACT – 0.573 x HSP70-hom (+2437)

Dividing each coefficient by 0.196 and rounding to the nearest whole number gives:

EBMT risk score + 2 x *MAL* rs8177374 – 2 x GCR haplotype ACT – 3 x HSP70-hom (+2437)

EBMT risk score can take values 0-7; *MAL* rs8177374 is coded CC = 0, TC or TT = 1; GCR haplotype ACT is coded 'GCR ACT' = 1, 'others' = 0; HSP70-hom (+2437) is coded TT = 0, CC or TC = 1.

Risk score elements are thus:

EBMT risk score:

values 0,1,2,3,4,5,6 or 7

MAL rs8177374 (patient):

CC: 0 x 2 = 0

TC or TT: 1 x 2 = 2

GCR haplotype ACT (patient):

'GCR ACT': 1 x -2 = -2 ; scaling linearly: -2 + 2 = 0

'others': 0 x -2 = 0; scaling linearly: 0 + 2 = 2

HSP70-hom (+2437) (patient):

TT: 0 x -3 = 0; scaling linearly: 0 + 3 = 3

CC or TC: 1 x -3 = -3; scaling linearly: -3 + 3 = 0

The clinical-genetic-score is derived after summing the aforementioned risk score elements of the EBMT risk score and genetic factors (final column, Table 2). For example, a patient with

EBMT risk score=3, presence of TC or TT for *MAL* in the patient (contribution 2), presence of ACT/ATA for "GCR ACT" in the patient (contribution 0) and presence of TT for HSP70-hom (+2437) in the patient (contribution 3) would have clinical-genetic-score= 3+2+0+3= 8.